

Tools and Technologies of Language Documentation
Prof. Bornini Lahiri and Prof. Dripta Piplai (Mondal)

Department of Humanities and Social Sciences

IIT Kharagpur

Week-02

Lecture-10

Lecture 10 : Designing Questionnaire, Part 1

Welcome to the 10th lecture of Tools and Technologies of Language Documentation. Today I will talk about the most important aspect of Language Documentation that is data collection. So, what are the methods through which we collect data? And how can we design questionnaires? There are various types of questionnaires, but today mainly we will talk about translation method. How we can collect data through translation method? How can we prepare questionnaire properly so that we can get more and proper data? What are the benefits of translation methods? And also, the limitations related to the translation method. So, we will talk about all these. So, let us begin with data elicitation.

In the very beginning of field work, the researcher has to rely on elicitation. What does elicitation mean? Elicitation means getting linguistic data from the native speakers by asking various types of questions. Generally, these are structured sessions. So, one plans the session once the appointments are taken from the language experts who can speak the language and then you have a questionnaire based on which you can ask different types of questions.

And one of the questions can be translating sentences or words that can be a part of the questionnaire and of course, then there can be different other types of questionnaires. So, this process is very easy to begin with because imagine you are in a field or you are among speakers who speak a very different language which you do not understand and then how do you actually collect data. So, of course, generally another language is used. So, there is a meta language through which you ask the questions and then the native speaker answers those in his or her language. So, to begin with this is easier because you do not know the language; slowly, you get hold of the language, you understand the words and sentence structures.

So, it is generally advised to start with this process. So, many field work manuals give

advice on what kind of questions needed to be asked in the field or what type of questions you should not ask in the field, because there are lots of restrictions as well. In the field you cannot offend your language experts, you cannot ask questions which can make the community feel bad. So, there are those type of restrictions. So in the field manuals, you will get lots of such instructions which you should follow while collecting data or data elicitation.

So, when we talk about methods, there are different types of methods which can be used for data collection. One and the most primary methods which is very common also is the translation method. And then of course, there are other methods like collecting narration, observation method, picture book method, video, object naming, discussion method, interview method, role playing and various other methods as well. But today, I will talk mainly about translation method. So in translation method what do we do? We create a set of words or sentences.

So we have a list prepared before going or meeting the language expert and then based on those sentences or words we ask them to translate it. So you can see here these are set of interrogative questions or interrogative sentences. So you can ask "What do you say what is your name in your language?", "How do you say where is your house in your own language?" So, you can ask and then you can get the answer and record it, write it in IPA, that is what is done. Sometimes if the native speaker can read the language, so, I have given here the example in English, but it can also be in any other language, which the native speaker understands. If he or she can read it, sometimes the list is also given them to read, they can read and reply.

But we do not ask the native speakers to write it. Why? That you can easily guess, because if they write, we will not get the exact pronunciation of the words. So generally what we do, we record it and also we can write in IPA at the same time, but of course, we need to check our IPA when we come back home, we listen to the recordings. So it is advised that the language speaker can read your sentences and say the exact translations. So, this is basically the translation method.

There can be sentences, there can be word list. You can even give them a story and then ask them to translate it, that can also be done. So, these are various methods; generally, when we do it, we make different sections of the translation sentences. So, like here you can see this is interrogative, there can be again a section based on case and agreement, there can be a section based on pronouns. So, there can be different types of sections made.

Similarly, for words also there are various sections made and then based on that,

language expert is asked to translate it. So, this is very easy method. Why words are important? So, I said sentences are translated and also words are taken. So, initially, generally we start with words, when we are working in a new language. First, to understand the phonology of the language because words in isolation can give us the phonological description of the language.

Of course, words in sentences are also required to look at its phonological description, but when we are new to a language, it is easier for us to understand the phonology if the words are in isolation. Because we know, due to various morphosyntactic rules phonology can change. So, for our own convenience or for the scholars' convenience, words in isolations is important actually. And then words in isolation can also help us create dictionaries, can help us create primers. So, these can also be used for the words which we collect.

There are various word list available. So which you can easily get in MPI tools, ICDS word list and the one, the most popular one is the Swadesh word list. So, Swadesh word list initially had 100 words, but now slowly words have increase in that, but generally people start with these basic 100 words. Why this is very important? Because in this Swadesh word list, what we see is collection of basic words and these words are common across languages mostly, not always of course, but these are taken as the words like I, you, they, eat, drink, because in most of the languages we find words for these type of actions or nouns or pronouns. These are the list which we find, these are the words which we find in the list of Swadesh.

So, Swadesh list is named after the linguist, Morris Swadesh. So, you can easily actually find Swadesh list online. Now some of the other questionnaires which are present are the Max Planck Institute's questionnaire, which is also very easily available. So, there is not only translation questionnaires, but also other questionnaires which I will talk later. So, elicitation with only one or with small groups that has been advised.

So, looking at the questionnaire, you can know how to elicit the data. There are certain exercises where there are fill in the blank. So, that requires knowledge of syntax by the language expert. So, there are also instructions which can be followed. But when you are adapting these questionnaires, it is important to say that you need to little bit change it, you need to customize it depending on the community where you are going to use the questionnaire.

There is SPPEL questionnaire, published by CIIL Publication, which you can explore and which is most suitable for Indian languages. There is another questionnaire, Jadavpur Journal of Languages and Linguistics, where you can find an issue on questionnaires

which is again prepared for Indian linguistic scenario. So, these are some of the questionnaires and in references you can find the details about these. So, there are some guidelines which are required whether we are adapting a questionnaire or we are preparing it from the scratch, in both the cases we need to keep some points in our mind. One is that we should always create some context while collecting sentences.

So, if you ask someone mango is sweet it is difficult to translate. While if you say the mango in the basket is sweet or the mango I gave you yesterday is sweet, then it is easier for one to translate. So, every time when we prepare translation questionnaire, it is very important that we give a certain context to the speaker. Sometimes we may not write that context in the questionnaire, we can only write a sentence, maybe we write the mango in the basket is sweet, but to the informant we give a context. We say him or we describe like imagine there is a basket full of mangoes and mango in the basket is sweet.

So, giving that background or context actually helps the translator to translate. You can actually imagine yourself in the place of the translator and then create easier way of translating the sentences. And the method should be made easier for the translator, so that he or she can give you more accurate data. Otherwise the language experts might feel bored, might feel tired easily and then you will not get proper data. Another very very important thing is that we need to customize our clues or target sentences.

So, what generally people do at times, people just read various questionnaires and as I said CIIL questionnaire or Jadavpur questionnaire, they are made for Indian public. They adopt it and they use it as it is in the field, but again, I would advise that those questionnaires should again be customized depending on the speech community. So, it is not like that that you have created one questionnaire and you can use it in all the different speech communities. So, it is always advised you can always customize it. So, what does customization mean? You should know little bit about the area where the language is spoken, you should have rough idea what are the vegetations, what is the geography of that area and then you can ask accordingly.

So, like if you give the sentence "Ram ate the apple" to a Toto speaker, he or she might not understand what you are saying or may understand, but will give you the word 'apple' for the word 'apple'. Because, they do not have any word for apple, apple is not part of their community, but instead if you say "Dhaniram ate the jackfruit", jackfruit is a very very common fruit almost everyone consumes it. So, it will be more convenient for them to translate the word jackfruit, they you will get a total word for jackfruit, but for apple you will get the same word apple. And the names can also be changed accordingly, the proper names. So, that the translator can feel more easy to translate, can identify the sentence easily.

So, that is always advised that you make the sentence like the community speakers, the way the speech community speaks the language, which vastly depends on the geography of the area. That is important. Again, another aspect mainly for collecting words is important is semantic domains. By semantic domain, what do we mean? Semantic domains are a particular domain where related meaning words are there.

So, semantically they are related. So, there can be different types of semantic domain like human beings and stages of life, physical environment. There can be animals, within animals you can have domestic animal, wild animal, you can have insects, reptiles. So, there can be different types of sub classifications. And why are these classifications or sub classifications required? Because, when we think, we think similar words together. So, if you ask someone what do you call a cockroach and then suddenly you ask what do you call a star, it might be little difficult to think.

Shifting from a cockroach to a star can be little difficult. But if I give you similar words, what do you call star, what do you call moon, what do you call the sun, then it is easier to translate or think; we think in that pattern. So cognitively, it is more convenient for us, when we think related meaning words. So, again as I mentioned, we should always try to design the questionnaire in such a way that it becomes easier for the translator to translate the words or sentences. So for the sake of convenience of the language experts, we should always create different types of semantic domains.

And as I mentioned, there can be one semantic domain within that, we can have sub classifications and again another sub classification. So, that can be done, but always related work should be together. So, if you are asking about body parts, you can ask about what do you call hand, what do you call fingers, what do you call nails and then you can slowly shift to like what do you call henna, which you do in your hands or what do you call.. you are talking about legs and you call what do you call the jewelry, which you wear in your legs or what do you call socks.

So, slowly there can be transition to maybe dress and ornaments. So, that transition should be smooth in one session. In another session of course, you can start with another semantic domain. But in the same session, if you want to continue, you can have a slow transition of the semantic domains. So semantic domain is very very important, and you can find lots of semantic domains and questions based on that over the internet.

You can easily search about it. So, what are the benefits of translation method? When you begin collecting data, you do not know about the language, it is a new language, then starting with the translation method is easier for the researcher. You have a planned

method, you have the questionnaire ready. You know that you are going to collect, say 100 sentences in one day, and then you get one-to-one match of the words and the sentences. You can easily know about the syntactic structure, the phonological part of the language.

So, these things become easier. And when you have less time, in that duration you can collect more data because the sessions are structured. You know that you have to collect like 100 or 200 sentences in 2 or 3 sittings. So, more data in less time. You get more data. Lots of time researchers, they do not have much time.

So, they can just rely on translation method. And it is also easier to clean and arrange the data, when you already have a set of organized data. You know that there are these 100 sentences and these are the data for those 100 sentences; it is already arranged or organized. You just need to come back home, you need to listen to the recordings and transcribe it and translate it. You need to delete the noises, the background noises and all, and that is it. You do not need to actually organize it from the scratch.

It is not like you do not know what was said or what was spoken, which can happen at times in narrations and other types of data collection. So, the benefits are that you can easily clean and arrange data. So, your data is actually ready in a short period of time. You go to field with a structured questionnaire, you sit for few sessions, you collect data, come back and get a very neat and clean data.

But then there are limitations to it. What are the limitations? First basic limitation is that you cannot get translation of the function words. So, those words which do not have a lexical meaning, but they are used in the sentences like of, for, about. So, these type of words are used in the language, they are part of the language. So, you cannot leave them right, you have to document them. So, when we are talking about documentation, we are actually trying to capture every aspect of a language and also metalinguistic aspects.

So, for other type of work like if I am collecting data for say, case and agreement or for numbers or for some other aspect, maybe I can depend only on translation method, but again, then also it is advised that you adapt other methods. But for documentation, only adapting translation method can never be enough because you will miss lots of other information. So, as I mentioned one is function words and then there can be certain words which are tabooed. So, it might not be tabooed in my language. So, I went and asked to the language expert and then I can find that the language expert got very angry because it was a tabooed word.

So, that can happen. Then in some languages, specially in many Tibeto-Burman

languages, we find words which cannot be translated or said in isolation like "village". So, it is always "my village" or "your village". Possession of the village, village cannot stand alone. Same stands for body parts, you cannot say "arm", "head", it has to be possessed by someone. So, it has to be like "my head", "my hand", something like that.

So, when you ask the language expert to give or translate these words in isolation, it becomes very difficult for the language expert. So you will not get accurate data. So, the language expert can ask "You are talking about whose village?". So in those cases, it is very very difficult to collect the words in isolation words like "house", "village". They hardly comes in isolation, they always need someone like "my house", "my village", words like that.

Again as I mentioned, about the function words, I will give you example from Swadesh 100 word list. I mentioned about Swadesh word list, which is very popular and widely used for language documentation or data collection. So, here you can see there are various words taken from the word list. And, in the first part, you can see there are words like "you".

So, these are pronouns "I, you, we, this, that". Now, when you talk about "you", and if you ask someone to translate the word "you", if someone ask me, I will be little confused because I have three forms of you, honorific, non-honorific. So, I do not know which one to give. That can be true in many communities. Similarly, words like "this" and "that".

So, maybe there are again other deictic elements. So, "this" and "that" and then "beyond that". So, words like that can be there. For "you" also, sometimes it can be inclusive, sometimes it could be exclusive. For "we" also, "we" pronoun can be sometimes inclusive, sometimes exclusive.

So, there can be actually variations of these pronouns. So, when we are looking at one-to-one translation, that might not be possible. Again, when you look at the second list, there are lots of function words like "at, in, with, if". So, if someone ask you, "What do you call "if" in your language?" or "What do you call "in" your language?", then that might be difficult to translate. So, these type of function words are actually very difficult to get translated. So, if you are preparing a translation questionnaire, then please avoid keeping these type of words in your questionnaire.

They are there in Swadesh list or other list, you might find it, but then they are very difficult to be translated and that is why they should be collected through some other method, but not translation method. Now again, another limitation is that when we are talking about language documentation, when we collect data through translation method,

we might not collect words which are reflecting local environments. Because, we have prepared our own questionnaire from our home and then we are in the field and then we are asking them to translate it. So, what about those flowers, trees or lots of other things which we can find only in that environment. May be there are no English or terms in any other language for those natural objects.

So, what can you do then? So, I talked about metalinguistic information, world view, all these are actually lost in translation method; these cannot be collected through translation method. The native speakers might at times feel embarrassed when you are asking to translate them some words. That can be because there is no equivalent of the word. So, maybe you are in say Toto Para, and you are forcing them to give word for say "apple". You are saying "No, there has to be a word in Toto for "apple", and give me the word".

That might not be the case. And then, there can be at times tabooed as I mentioned. And sometimes, in cases of endangered languages, speakers might feel embarrassed because they have forgotten their language. They cannot recall what is the word in their language. And then forcing them to translate it, might lead to embarrassment for them, also they can feel uncomfortable.

And so we should avoid those situations. We should always make our language experts, very comfortable, so that, we can get data from the person and the person should not feel actually embarrassed or bad. He or she should also enjoy the whole process. So, when I talked about the local environments, you can see this example from Kurmali. So, what we found in Kurmali was that, there are lots of words for say "pond, small pond, river, small river", but since we collected this data from Purulia, which is a very dry area, there is also a word for "river, which has no water in summer". So in summer, the water dries out from this area and that is why, they have a name for that type of landscape, which is a river in other season, but in summer it is dried out.

So now, if we have a prepared translation method questionnaire, we might have words like "pond, river" in that, but it is very unlikely to have a word for a "river, which has no water". So that, we can only know when we are in that locality and then again, we cannot collect it through translation method, but through some other method. So, I would like to conclude by saying there are various methods of linguistic data collection. Of course, depends on what is your work or what is your research area and depending on that, one can choose the particular method. When we talk about language documentation, of course it is good to begin with translation method, but then there are other methods which needs to be adapted.

So, translation method is most convenient and that is why it is the most popular one and

also it is very easier on the researchers part to prepare a questionnaire and then go to the field and ask them to translate. But before preparing a questionnaire for translation method, we should know that area where the language is spoken. We should have a little idea about the speech community and that is why we do little pilot surveys. So sometimes, one or two days pilot surveys are done, where you actually survey the landscape. You know little bit about the speech community, which helps you in creating your own questionnaire or customizing a questionnaire.

Nowadays, we can also browse a little about the area, we can do a little search online and know about the area and the speech community. So that, we can make the questionnaire more convenient for the speech community. So while preparing questionnaire, we should be aware of the landscape and we should customize it according to the speech community. This prior field knowledge helps us to actually create the questionnaire better for the data elicitation. Translation method is one of the most popular methods of data collection for its convenience and as I mentioned, in little time, you can collect more data, but then there are some major limitations to this process.

One is that we cannot capture everything through translation method; lots of linguistic part cannot be collected through it and then the metalinguistic information of course, cannot be collected using translation method. You can also miss on the local terms, which is related to local geography, local area. So, all that can be missed in the translation method, but do not worry. We have also other methods to collect data for language documentation, which I will talk about in my next lecture. So, these are some of the references for different types of questionnaires and most of them are easily and freely downloadable.

So, you can actually explore. You can also find which one is more suitable for Indian languages. You can look at these questionnaire and think about the limitations, what you find in these questionnaires. So, I hope you enjoyed this class. Thank you!